



Genome Assembly of *Azotobacter chroococcum* Strain W5, a Free-Living Diazotroph Isolated from India

Annapurna Kannepalli,^a Kedharnath Reddy Pengani,^a Jeshima Khan Yasin,^b Sangeeta Paul,^a Swati Tyagi,^a Govindasamy Venkadasamy,^a Meenakshi Sharma,^a Swarnalakshmi Karivaradharajan^a

^aDivision of Microbiology, ICAR-Indian Agricultural Research Institute, New Delhi, India

^bDivision of Genomic Resources, ICAR-NBPGR, PUSA Campus, New Delhi, India

ABSTRACT *Azotobacter chroococcum* strain W5 (MTCC 25045) is an effective diazotrophic bacterium with plant growth-promoting traits. Here, we report the draft genome assembly of this biologically and agronomically evaluated *A. chroococcum* strain. The genome assembly in 55 contigs is 4,617,864 bp long, with a G+C content of 66.83%.

A *zotobacter* spp. are Gram-negative, free-living, aerobic, soil-dwelling, oval or spherical bacteria that form thick-walled cysts (as a means of asexual reproduction under favorable conditions). There are seven known species of the genus.

A Gram-negative, aerobic, nitrogen-fixing bacterium was isolated from the rhizosphere of wheat (*Triticum aestivum* L.) grown at Indian Agricultural Research Institute (IARI) research farms and identified as *Azotobacter chroococcum* based on 16S rRNA gene sequencing (GenBank accession no. [MT299751](#)). The bacterium grew luxuriously on N₂-free mineral medium and produced brown pigment and plant growth-promoting hormones (1). This isolate was named *Azotobacter chroococcum* strain W5 (MTCC 25045). Inoculation with *A. chroococcum* strain W5 led to an increase in crop yield (2, 3). These findings prompted further efforts to investigate this bacterium for potential applications in agricultural and pharmaceutical industries.

To identify the specific genomic features of this special strain, we attempted whole-genome sequencing. Genomic DNA (gDNA) was extracted as reported earlier (4) from a 72-h culture grown in Luria-Bertani medium that was inoculated with a single isolated colony using a standard isolation protocol as per the manufacturer's instructions (Qiagen DNA extraction kit). The integrity and quality of the DNA were checked through a fluorescence-based quantitation assay as per the developer's instructions (Qubit 3.0; Thermo Fisher). An ultrasonicator (Covaris) was used to shear the genomic DNA and to generate 350-bp-long fragments. DNA libraries (NEBNext Ultra DNA library prep kit; Illumina, USA) were generated following the manufacturer's recommendations. The fragments were end polished, A tailed, and ligated to full-length adapters for Illumina sequencing. In the final preparation step, PCR products were purified (AMPure XP system). Libraries were analyzed for size distribution using a Bioanalyzer (Agilent 2100) (5) and quantified using real-time PCR (CFX96; Bio-Rad). Illumina paired-end sequencing (HiSeq 1500) resulted in 14,040,301 raw reads with a read length of 150 bases, totaling 4.2 GB of data output. The quality of the reads generated was assessed using FastQC v0.11.5 (6) (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Of the total raw reads generated, 14.24% were below a Phred score of 40. Raw reads with N (unidentified base) and Phred scores less than 30 were filtered out using Trimmomatic v0.39 (7, 8) before further analyses with Illumina adapter sequences (ATCACG) (9). The high-quality clean data of 3.6 Gb with 13,178,049 reads were sub-

Citation Kannepalli A, Pengani KR, Yasin JK, Paul S, Tyagi S, Venkadasamy G, Sharma M, Karivaradharajan S. 2020. Genome assembly of *Azotobacter chroococcum* strain W5, a free-living diazotroph isolated from India. *Microbiol Resour Announc* 9:e00259-20. <https://doi.org/10.1128/MRA.00259-20>.

Editor Christina A. Cuomo, Broad Institute

Copyright © 2020 Kannepalli et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](#).

Address correspondence to Annapurna Kannepalli, annapurna96@yahoo.co.in.

Received 20 March 2020

Accepted 23 April 2020

Published 14 May 2020

TABLE 1 *Azotobacter chroococcum* strain W5 sequencing and assembly metrics

Genome feature	Value(s)
Sequence length (bp)	4,617,864
No. of contigs	55
G+C content (%)	66.83
Shortest contig length (bp)	248
Longest contig length (bp)	476,631
No. of repeat regions	1
Total no. of genes	4,335
Total no. of CDSs ^a	4,273
Total no. of coding genes	4,168
No. of tRNAs	53
No. of rRNAs	5 (3 5S, 1 16S, and 1 23S)
No. of ncRNAs	4
No. of CRISPRs	1
N_{50} (bp)	257,772
Total no. of pseudogenes	105

^aCDSs, coding DNA sequences.

jected to genome assembly using SPAdes v3.11.1 with default settings (10). Table 1 displays the assembly details of *Azotobacter chroococcum* strain W5.

The smallest and largest contig sizes were 248 and 476,631 bases, respectively. Finally, 55 contigs with an N_{50} value of 257,772 bases were assembled into the genome of 4,617,864 bases, coding 4,335 genes and 105 pseudogenes. Furthermore, 53 tRNAs, 5 rRNAs, and 4 noncoding RNAs (ncRNAs) were predicted by using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) v4.11 (11). Overall, the G+C content of the assembly was 66.83%. Siderophore- and bacteriocin-coding genes were predicted by antiSMASH v5.0.0 using default settings (12).

Data availability. The complete genome sequence of *A. chroococcum* strain W5 has been deposited in DDBJ/ENA/GenBank under the accession no. JAAPAP000000000 (BioProject no. PRJNA610299, BioSample no. SAMN14292004, and SRA accession no. SRR11237782). The version described in this paper is the first version, JAAPAP010000000.

ACKNOWLEDGMENTS

This work was supported in part by funding received from the Indian Council of Agricultural Research-Incentivizing Research in Agriculture on Biological Nitrogen Fixation (ICAR-IRA BNF) and the Indo-UK IUNFCproject of the Department of Biotechnology (BT/IN/UK/VNC-41/DLN/2015-16), Government of India, New Delhi, India.

The facility provided by the director of ICAR-IARI (New Delhi, India) is gratefully acknowledged.

We declare no conflict of interest.

REFERENCES

- Apte R, Shende ST. 1981. Studies on *Azotobacter chroococcum* I. Morphological, biochemical and physiological characterization of *A. chroococcum*. Zentralbl Bakteriol B 136:548–554. [https://doi.org/10.1016/S0323-6056\(81\)80066-3](https://doi.org/10.1016/S0323-6056(81)80066-3).
- Pandey A, Shende ST. 1991. Effect of *Azotobacter chroococcum* inoculation on yield and post harvest seed quality of wheat (*Triticum aestivum*). Zentralbl Mikrobiol 146:489–494. [https://doi.org/10.1016/S0232-4393\(11\)80236-2](https://doi.org/10.1016/S0232-4393(11)80236-2).
- Malik BS, Paul S, Sharma RK, Sethi AP, Verma OP. 2005. Effect of *Azotobacter chroococcum* on wheat (*Triticum aestivum*) yield and its attributing components. Ind J Agri Sci 75:600–602.
- Annapurna K, Govindasamy V, Sharma M, Rajrana Y, Swarnalakshmi K, Paul S, Ghosh A, Chikara SK. 2017. Draft genome sequence of *Pseudomonas stutzeri* strain KMS 55, an endophytic diazotroph isolated from rice roots. Genome Announc 5:e00972-17. <https://doi.org/10.1128/genomeA.00972-17>.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
- Duong DA, Espinosa-Artiles P, Orozco RA, Molnár I, Stock SP. 2019. Draft genome assembly of the entomopathogenic bacterium *Photorhabdus luminescens* subsp. sonorensis Caborca. Microbiol Resour Announc 8:e00692-19. <https://doi.org/10.1128/MRA.00692-19>.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina Sequence Data. Bioinformatics 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
- Robson RL, Jones R, Robson RM, Schwartz A, Richardson TH. 2015. *Azotobacter* genomes: the genome of *Azotobacter chroococcum* NCIMB 8003 (ATCC 4412). PLoS One 10:e0127997. <https://doi.org/10.1371/journal.pone.0127997>.
- Annapurna K, Govindasamy V, Sharma M, Ghosh A, Surendra K. 2018. Whole genome shotgun sequence of *Bacillus paralicheniformis* strain

- KMS 80, a rhizobacterial endophyte isolated from rice (*Oryza sativa* L.). 3 Biotech 8:223. <https://doi.org/10.1007/s13205-018-1242-y>.
10. Tran PN, Tan NEH, Lee YP, Gan HM, Polter SJ, Dailey LK, Hudson AO, Savka MA. 2015. Whole-genome sequence and classification of 11 endophytic bacteria from poison ivy (*Toxicodendron radicans*). Genome Announc 3:e01319-15. <https://doi.org/10.1128/genomeA.01319-15>.
 11. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. Nucleic Acids Res 44:6614–6624. <https://doi.org/10.1093/nar/gkw569>.
 12. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T. 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res 47:W81–W87. <https://doi.org/10.1093/nar/gkz310>.